DOCUMENT RESUME

ED 438 801 IR 019 929

AUTHOR Choo, Chun Wei; Detlor, Brian; Turnbull, Don
TITLE Information Seeking on the Web--An Integrated Model of Browsing and Searching.
PUB DATE 1999-11-00
NOTE 16p.; Paper presented at the Annual Meeting of the American Society for Information Science (ASIS) (62nd, Washington, D.C., October 31-November 4, 1999).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Behavioral Science Research; Foreign Countries; Information Scientists; *Information Seeking; Internet; Media Specialists; Models; *Online Searching; *Search Strategies; User Needs (Information); User Satisfaction (Information); World Wide Web
IDENTIFIERS Browsing

ABSTRACT

This paper presents findings from a study of how knowledge workers use the Web to seek external information as a part of their daily work. Participants were mainly IT specialists, managers, and research/marketing/consulting staff working in organizations that included a large utility company, a major bank, and a consulting firm. Thirty-four participants answered a detailed questionnaire and were interviewed individually in order to understand their information needs and information seeking preferences. A custom-developed WebTracker software application installed on the participants' workplace personal computers continuously recorded participants' activities during two-week periods. Sixty-one significant episodes of information seeking were identified. A model, based on previous research, was developed to describe the common repertoires of information seeking. The study suggests that a behavioral framework, which relates motivations and moves, may be helpful in analyzing patterns of Web-based information seeking. (AEF)

Reproductions supplied by EDRS are the best that can be made
from the original document.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
■ This document has been reproduced as received from the person or organization originating it.
□ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

B. Detlor

1999 ASIS Annual Meeting Contributed Paper

INFORMATION SEEKING ON THE WEB – AN INTEGRATED MODEL OF BROWSING AND SEARCHING

Chun Wei Choo, Brian Detlor, Don Turnbull
(choo@fis.utoronto.ca, detlor@fis.utoronto.ca, turnbull@fis.utoronto.ca)
Faculty of Information Studies, University of Toronto, Toronto, Ontario, Canada

ED 438 801

Abstract

The paper presents findings from a study of how knowledge workers use the Web to seek external information as part of their daily work. Thirty four users from seven companies took part in the study. Participants were mainly IT specialists, managers, and research/marketing/consulting staff working in organizations that included a large utility company, a major bank, and a consulting firm. Participants answered a detailed questionnaire and were interviewed individually in order to understand their information needs and information seeking preferences. A custom-developed WebTracker software application was installed on each of their workplace PCs, and participants' Web-use activities were then recorded continuously during two-week periods. The WebTracker recorded how participants used the browser to seek information on the Web: it logged menu choices, button bar selections, and keystroke actions, allowing browsing and searching sequences to be reconstructed. In a second round of personal interviews, participants recalled critical incidents of using information from the Web.

Data from the two interviews and the WebTracker logs constituted the database for analysis. Sixty one significant episodes of information seeking were identified. A model was developed to describe the common repertoires of information seeking that were observed. On one axis of the model, episodes were plotted according to the four scanning modes identified by Aguilar (1967), Weick and Daft (1983): undirected viewing, conditioned viewing, informal search, and formal search. Each mode is characterized by its own information needs and information seeking strategies. On the other axis of the model, episodes were plotted according to the occurence of one or more of the six categories of information seeking behaviors identified by Ellis (1989, 1990): starting, chaining, browsing, differentiating, monitoring, and extracting. The study suggests that a behavioral framework that relates motivations (Aguilar) and moves (Ellis) may be helpful in analysing patterns of Web-based information seeking.

1 RESEARCH OBJECTIVES

The research presented in this paper has three objectives:

(1) To develop a behavioral model of information seeking on the Web based on modes of browsing and searching differentiated by information needs and information seeking activity;

(2) To develop operational methods for measuring information seeking on the Web by analyzing browser-based actions and events;

(3) To combine the use of multiple, complementary methods of collecting qualitative and quantitative data on how people seek and use Web-based information in their natural work settings.

The paper is organized into five sections. Section 2 introduces recent research on information seeking on the Web. Section 3 integrates elements from research in information seeking and organizational scanning

IR019929

BEST COPY AVAILABLE

into a behavioral model of Web-based information seeking. Section 4 presents results from our study which appear to be compatible with the proposed model. Section 5 is a summary.

2 Recent Research on Information Seeking on the Web

Until recently, there were few direct, rigorous studies of Web browsing behavior despite the Web's growing popularity. One reason is the difficulty in collecting complete sets of data to describe Web browsing sessions. To obtain data on Web information seeking, Web use logs should preferably be collected on the Web browsing client system. Web or Proxy server logs provide excellent volume or Web usage, but they do not capture Web access from the browser's local cache, which typically provides most of the Web pages requested via the Back and Forward buttons in Web browsers. Other browser actions that are not logged include bookmarking, printing a Web page, or finding terms in an open page.

Catledge and Pitkow (1995) were the first to publish a major study of Web browsing behavior by modifying the source code for a version of XMosaic, the dominant X Windows browser at the time. They configured the browser to generate a client-side log file that showed user navigation strategies and interface selections. They released this modified browser to Computer Science department students who ran Mosaic from X Terminals in the various departmental computing labs at Georgia Tech. Results were measured using a task-oriented method. They determined session boundaries by analyzing the time between each event for all events, and adopted the heuristic that a lapse of 25.5 minutes or greater indicated the end of a "session." This heuristic is currently the most-commonly used for delimiting sessions. The study yielded some unexpected results. Web pages that users bookmarked did not match the most-popular sites visited as a whole from the group. Only 2% of Web pages were either saved locally or printed. These results may have been influenced by limitations in the browser (XMosaic's bookmarking capabilities), or the availability of printers in the work place. Catledge and Pitkow also hypothesized that users in their study categorized as "browsers" spend less time on a Web page than "searchers."

Tauscher and Greenberg (1997a, 1997b) focused on the history mechanisms that Web browsers use to manage recently-requested Web pages browsed in a session of Web information seeking. They also used a modified XMosaic browser to collect Web browsing data for over six weeks from 23 participants. They recorded and examined the rate that Web pages were visited; how users visited old and new Web pages; the distance (in terms of URLs) between repeated Web page visits; the frequency of Web page visits, the extent of browsing in one cluster of Web pages; and repeated sequences of "path-following behavior." (1997a, p. 400) Most significantly, they found that 58% of the pages visited during a Web browsing session were re-visits. This seems to suggest that Web information seeking may be influenced by Web browser functionality that makes it easy to go back to recently viewed pages. Tauscher and Greenberg showed that overall, users also only access a few pages frequently (60% once, and 19% twice) and browse in very small clusters of pages. They contend that Web browsing activity is a "*recurring system*... where users predominantly repeat activities they had invoked before, while still selecting new actions from the many that are possible." (1997a, p. 400) People explained that they revisited Web pages because "the information contained by them changes; they wish to explore the page further; the page has a special purpose (e.g. search engine, home page); they are authoring a page; or the page is on a path to another revisited page." (1997a, p. 400). Thus, Tauscher and Greenberg identified seven Web browsing patterns: first-time visits to a cluster of pages; revisits to pages; page authoring (where the subject used Reload to view the newly modified page); use of web-based applications; hub-and-spoke visits (navigating to each new page from around a central page); a guided tour where links guide navigation through the Web pages; and a depth-first search where link paths are followed without returning to the first page in some cases.

More recently, Huberman, Pirolli, Pitkow and Lukose (1998) discovered several strong regularities of Web user surfing patterns, and developed a mathematical "law of surfing, ... that determines the probability distribution of the depth – that is, the number of page a user visits within a Web site." (p. 95) They started with a model of probability of the number of links a user might follow on a Web site. Next they calculated a value for the current page and related this value to the next page accessed that leads to examining the cost of continuing surfing. When the cost of moving to the next Web page is more than its expected value, the user stops Web surfing. They analyzed data collected from a sample of AOL (America Online) users for each of five days, a huge amount of data. One day alone (December 5, 1997) yielded 23,692 AOL users who collectively surfed 3,247,054 Web pages from 1,090,168 unique Web pages. This amount of data is staggering compared to previous studies of Web use.

In a related study, Huberman et. al. examined Web server logs of the Xerox external Web site in order to obtain a constrained set of Web page requests. They used "cookies" to help track the paths of individual users as they surfed through the Web site. Generally, they found a "strong fit" which was consistent through each day of the study. By applying this model along with a spreading activation algorithm, they could predict the number of requests for each Web page in a Web site. As they point out, this has implications for e-commerce applications and Web site organization, not to mention providing a more robust understanding of information seeking patterns on the Web. Overall, their study echoes other research in suggesting that "surfing patterns on the Web display strong statistical regularities that can be described by a universal law. In addition, the success of the model points to the existence of utility maximizing behavior underlying surfing." (p.97) These findings do not signal the end of new findings about Web information seeking, but do establish a firm foundation to build upon in further research.

3 Towards a New Behavioral Model of Information Seeking on the Web

3.1 Modes of Browsing and Searching

Marchionini (1995) reviewed the research on browsing and observed that "there seems to be agreement on three general types of browsing that may be differentiated by the object of search (the information needed) and by the systematicity of tactics used." (p. 106) **Directed browsing** occurs when browsing is systematic, focused, and directed by a specific object or target. Examples include scanning a list for a known item, and verifying information such as dates or other attributes. **Semidirected browsing** occurs when browsing is predictive or generally purposeful: the target is less definite and browsing is less systematic. An example is entering a single, general term into a database and casually examining the retrieved records. Finally, **undirected browsing** occurs when there is no real goal and very little focus. Examples include flipping through a magazine and "channel-surfing."

In a similar vein, Wilson (1997) identifies the following categories of information seeking and acquisition after a survey of research that included health information seeking.

> **Passive attention**: such as listening to the radio or watching television programmes, where there may be no information-seeking intended, but where information acquisition may take place nevertheless;
>
> **Passive search**: which seems like a contradiction in terms, but signifies those occasions when one type of search (or other behavour) results in the acquisition of information that happens to be relevant to the individual;
>
> **Active search**: which is the type of search most commonly thought of in the information science literature, where an individual actively seeks out information; and

Ongoing search: where active searching has already established the basic framework of ideas, beliefs, values, or whatever, but where occasional continuing search is carried out to update or expand one's framework.

It is interesting to observe that in a separate stream of research in organization science, a comparable categorization of modes of organizational scanning or "browsing" has been proposed, based on both empirical and theoretical research. The initial field work of Aguilar (1967) and the subsequent theoretical expansion by Weick and Daft (Weick and Daft 1983; Daft and Weick 1984) suggest that organizations scan in four distinct modes: undirected viewing, conditioned viewing, informal search, formal search. In this study, we amplify the information seeking implications of each of these modes, by elaborating on how directed the scanning would be, and on the amount and kind of effort expended (Figure 1). (The modes of viewing presented here are comparable and compatible with the three general types of browsing that Marchionini (1995) identified. However, because we use "browsing" in the next section to describe a pattern of micro-moves, we retain the term "viewing" here to avoid confusion and to indicate provenance.)

In **undirected viewing**, the individual is exposed to information with no specific informational need in mind. The overall purpose is to scan broadly in order to detect signals of change early. Many and varied sources of information are used, and large amounts of information are screened. The granularity of information is coarse, but large chunks of information are quickly dropped from attention. The goal of broad scanning implies the use of a large number of different sources and different types of sources.

In **conditioned viewing**, the individual directs viewing to information about selected topics or to certain types of information. The overall purpose is to evaluate the significance of the information encountered in order to assess the general nature of the impact on the organization. The individual has isolated a number of areas of potential concer n from undirected viewing, and is now sensitized to assess the significance of developments in those areas.

During **informal search**, the individual actively looks for information to deepen the knowledge and understanding of a specific issue. It is informal in that it involves a relatively limited and unstructured effort. The overall purpose is to gather information to elaborate an issue so as to determine the need for action by the organization.

During **formal search**, the individual makes a deliberate or planned effort to obtain specific information or types of information about a particular issue. Search is formal because it is structured according to some pre-established procedure or methodology. The granularity of information is fine, as search is relatively focused to find detailed information. The overall purpose is to systematically retrieve information relevant to an issue in order to provide a basis for developing a decision or course of action. The four modes of scanning are summarized and compared in Figure 1.

Scanning Modes	Information Need	Information Seeking	Information Use
Undirected Viewing	General areas of interest; specific need to be revealed	• Scan broadly a diversity of sources, taking advantage of what's easily accessible • "**Sweeping**"	Serendipitous discovery "**Browsing**"
Conditioned Viewing	Able to recognize topics of interest	• Browse in pre-selected sources on pre-specified topics of interest • "**Discriminating**"	Increase knowledge about topics of interest "**Learning**"
Informal Search	Able to formulate simple queries	• Search is focused on area or topic, but a good-enough search is satisfactory • "**Satisficing**"	Increase knowledge on area within narrow boundaries "**Selecting**"
Formal Search	Able to specify targets in detail	• Systematic gathering of information about an entity, following some method or procedure • "**Optimizing**"	Formal use of information for decision-, policy-making "**Retrieving**"

Figure 1. Modes of Scanning

3.2 Ellis' Model of Information Seeking Behaviors

Ellis (1989), Ellis et al (1993), and Ellis and Haugan (1997) propose and elaborate a general model of information seeking behaviors based on studies of the information seeking patterns of social scientists, research physicists and chemists, and engineers and research scientists in an industrial firm. One version of the model describes six categories of information seeking activities as generic: starting, chaining, browsing, differentiating, monitoring, and extracting.

Starting comprises those activities that form the initial search for information – identifying sources of interest that could serve as starting points of the search. Identified sources often include familiar sources that have been used before as well as less familiar sources that are expected to provide relevant information. While searching the initial sources, these sources are likely to point to, suggest, or recommend additional sources or references. Following up on these new leads from an initial source is the activity of **Chaining**. Chaining can be backward or forward. Backward chaining takes place when pointers or references from an initial source are followed, and is a well established routine of information seeking among scientists and researchers. In the reverse direction, forward chaining identifies and follows up on other sources that refer to an intial source or document. Although it can be an effective way of broadening a search, forward chaining is much less commonly used.

Having located sources and documents, **Browsing** is the activity of semi-directed search in areas of potential search. The individual often simplifies browsing by looking through tables of contents, lists of titles, subject headings, names of organizations or persons, abstracts and summaries, and so on. Browsing takes place in many situations in which related information has been grouped together according to subject affinity, as when the user views displays at an exhibition, or scans books on a shelf. ("Browsing" in Ellis' model is different from "viewing" in the previous section: browsing here describes looking for information at the micro-event level; whereas viewing earlier describes a broader context of looking at information.)

During **Differentiating**, the individual filters and selects from among the sources scanned by noticing differences between the nature and quality of the information offered. For example, social scientists were found to prioritize sources and types of sources according to three main criteria: by substantive topic; by approach or perspective; and by level, quality, or type of treatment (Ellis 1989). The differentiation process is likely to depend on the individual's prior or initial experiences with the sources, word-of-mouth recommendations from personal contacts, or reviews in published sources.

Monitoring is the activity of keeping abreast of developments in an area by regularly following particular sources. The individual monitors by concentrating on a small number of what are perceived to be core sources. Core sources vary between professional groups, but usually include both key personal contacts and publications. **Extracting** is the activity of systematically working through a particular source or sources in order to identify material of interest. As a form of retrospective searching, extracting may be achieved by directly consulting the source, or by indirectly looking through bibliographies, indexes, or online databases. Retrospective searching tends to be labor intensive, and is more likely when there is a need for comprehensive or historical information on a topic.

Marchionini (1995) proposes another often-cited model of the information-seeking process, tuned perhaps to electronic environments. In his model, the information seeking process is composed of eight subprocesses which develop in parallel: (1) recognize and accept an information problem, (2) define and understand the problem, (3) choose a search system, (4) formulate a query, (5) execute search, (6) examine results, (7) extract information, and (8) reflect/iterate/stop. (Marchionini 1995, p. 49-60). The subprocess of "extract information" bears the same name as Ellis' "extracting" activity but the two processes are different. Marchionini (1995) describes extracting thus: "There is an inextricable relationship between judging information to be relevant and extracting it for all or part of the problem's solution. ... To extract information, an information seeker applies skills such as reading, scanning, listening, classifying, copying, and storing information. ... As information is extracted, it is manipulated and integrated into the information seeker's knowledge of the domain." (p. 57-58) In Ellis' model, "browsing" and "differentiating" are activities separate from "extracting," which is "systematically working through a particular source or sources to identify material of interest." (Ellis 1989, p. 242) On the Web, we expect extracting (in Ellis' sense) to mean systematically working through a selected website or set of web pages (typically using search engines) in order to search and retrieve material of interest.

Ellis (1989) thought that hypertext-based systems would have the capabilities to implement functions indicated by his behavioral model. If we visualize the World Wide Web as a hyperlinked information system distributed over numerous networks, most of the information seeking behavior categories in Ellis' model are already being supported by capabilities available in common Web browser software. Thus, an individual could begin surfing the Web from one of a few favourite starting pages or sites (starting); follow hypertextual links to related information resources – in both backward and forward linking directions (chaining); scan the Web pages of the sources selected (browsing); bookmark useful sources for future reference and visits (differentiating); subscribe to e-mail based services that alert the user of new information or developments (monitoring); and search a particular source or site for all information on that site on a particular topic (extracting). Plausible extensions of the acitivities to Web information seeking (labelled *Web Moves*), are compared with the original formulations (*Literature Search Moves*) in Figure 2 below.

	Starting	Chaining	Browsing	Differentiating	Monitoring	Extracting
Literature Search Moves (Ellis et al, 1989, 1993, 1997)	Identifying sources of interest	Following up references found in given material	Scanning tables of contents or headings	Assessing or restricting information according to their usefulness	Receiving regular reports or summaries from selected sources	Systematically working a source to identify material of interest
Anticipated Web Moves	Identifying websites/ pages containing or pointing to information of interest	Following links on starting pages to other content-related sites	Scanning top-level pages: lists, headings, site maps	• Selecting useful pages and sites by bookmarking, printing, copying and pasting, etc. • Choosing differentiated, pre-selected site	• Receiving site updates using push, agents, or profiles • Revisiting 'favorite' sites	Systematically searches a local site to extract information of interest at that site

Figure 2. Information Seeking Behaviors and Web Moves

3.3 Towards a Behavioral Model of Information Seeking on the Web

Aguilar's modes of scanning and Ellis's seeking behaviors may be combined and extended in a new behavioral model of information seeking on the Web. The figure below identifies four main modes of information seeking on the Web: undirected viewing, conditioned viewing, informal search, and formal search. For each mode, the figure indicates which information seeking activities or moves are likely to occur frequently, as suggested by theory.

	Starting	Chaining	Browsing	Differentiating	Monitoring	Extracting
Undirected Viewing	Identifying selecting starting pages, sites	Following links on initial pages				
Conditioned Viewing			Browsing entry pages, headings, site maps	Bookmarking, printing, copying Going directly to known site	Revisiting 'favorite' or bookmarkedsites for new information	
Informal Search				Bookmarking, printing, copying Going directly to known site	Revisiting 'favorite' or bookmarkedsites for new information	Using (local) search engines to extract information
Formal Search					Revisiting 'favorite' or bookmarkedsites for new information	Using search engines to extract information

Figure 3. Behavioral Modes and Moves of Information Seeking on the Web

3.3.1 Undirected Viewing

In the undirected viewing mode on the Web, we expect to see many instances of **starting** and **chaining**. Starting occurs when viewers begin their web use on pre-selected default home pages, or when they visit a favorite page or site to begin their viewing (such as news, newspaper, or magazine sites). Chaining occurs when viewers notice items of interest (often by chance), and then follow hypertext links to more information on those items. Forward chaining of the sort just described is the most typical during undirected viewing. Backward chaining is also possible, since search engines can be used to locate other Web pages that point to the site that the user is currently at.

3.3.2 Conditioned Viewing

In the conditioned viewing mode on the Web, we expect **browsing**, **differentiating**, and **monitoring** to be common. Differentiating occurs as viewers select Web sites or pages that they expect to provide relevant information. Sites may be differentiated based on prior personal visits, or recommendations by others (such as word-of-mouth or published reviews). Differentiated sites are often bookmarked. When visiting differentiated sites, viewers browse the content by looking through tables of contents, site maps, or list of items and categories. Viewers may also monitor highly differentiated sites by returning regularly to browse, or by keeping abreast of new content (through, for example subscribing to newsletters that report new material on the site).

3.3.3 Informal Search

During informal search on the Web, we expect **differentiating**, **extracting**, and **monitoring** to be typical. Again, informal search is likely to be attempted at a small number of Web sites that have been differentiated by the individual, based on the individual's knowledge about these sites' information relevance, quality, affiliation, dependability, and so on. Extracting is relatively "informal" in the sense that searching would be localized to looking for information within the selected site(s). Extracting is also likely to make use of the basic, 'simple' search features or commands of the local search engine, in order to get at the most important or most recent information, without attempting to be comprehensive. Monitoring becomes more proactive if the individual sets up push channels or software agents that automatically find and deliver information based on keywords or subject headings.

3.3.4 Formal Search

During formal search on the Web, we expect primarily **extracting** operations, with some complementary **monitoring** activity. Formal search makes use of search engines that cover the Web relatively comprehensively, and that provide a powerful set of search features that can focus retrieval. Because the individual wishes not to miss any important information, there is a willingness to spend more time in the search, to learn and use complex search features, and to evaluate the sources that are found in terms of quality or accuracy. Formal search may be two-staged: multi-site searching that identifies significant sources is then followed by within-site searching. Within-site searching may involve fairly intensive foraging. Extracting may be supported by monitoring activity, again through services such as Web site alerts, push channels/agents, and e-mail announcements, in order to keep up with late-breaking information.

4 Research Design and Methods

4.1 Participants

Thirty four participants from seven companies took part in the study. Since participants who regularly use the Web as part of their daily work were preferred, volunteers were canvassed through invitations at various IT-related workshops and conferences; postings at technology-focused listservs; and direct e-mail contact with colleagues and associates at large technology-oriented companies.

The seven companies comprised a major national bank, a large utility company, a national magazine publisher; a medium-sized University research library, a medium-sized marketing agency, and two small software consulting firms. The participants held jobs as IT technical specialists or analysts; managers; researchers; marketing staff; consultants; and administrative staff.

All of the users in this study primarily utilized the Web for business purposes as an integral part of their work responsibilities and activities. In most cases, participants were connected to the Internet through continuous leased-line access and used relatively high-powered machines. Many of the participants would be generally regarded as technically proficient Web users.

4.2 Data Collection

Three methods of data collection were employed: questionnaire survey; a WebTracker application that recorded Web browser actions; and personal interviews with participants. [A more detailed description of the data collection procedure is in Choo, Detlor, and Turnbull (1998).]

The **questionnaire survey** was administered at the participants' work places, during the first site visit. The survey contained 12 questions that identifed the information sources the participants used, their frequency of using these sources, and their perception of the perceived accessibility and quality of each of the sources. A wide range of sources was covered, including personal and impersonal sources (print and electronic), as well as internal and external sources. There were also questions on the amount of time and frequency of using the Web for information seeking. Furthermore, through informal conversations during the visit, research team members were able to develop a general impression of the style and scope of each participant's Web use.

The custom-developed **WebTracker application** was installed on each participant's computer, and it ran transparently whenever the participant's Web browser was being used. The WebTracker application was left to run on participants' computers for two-week periods. Because the WebTracker was essentially 'invisible,' it was not expected to influence participants' normal Web-use behaviors. After two weeks, WebTracker was uninstalled, and the WebTracker log file collected for analysis. The WebTracker recorded how each participant was using the browser to navigate the Web and manipulate information from the Web. Specifically, it recorded all URL calls and requests, as well as most browser menu selections, and wrote these events into a local log file on each participant's hard disk. Browser menu selections captured included "Open URL or File," "Reload," "Back," "Forward," "Add to Bookmarks," "Go to Bookmark," "Print," and "Stop." Because all URL calls and menu selections were date-time stamped as they were written into the WebTracker log, the research team was able to subsequently reconstruct move-by-move how participants looked for information on the Web during particular episodes.

The WebTracker log was pre-analyzed to prepare for **personal interviews** with each participant. The interview format was based on the principles of the Critical Incident Technique (Flanagan 1954), in which the 'incident' to be studied should be recent, sufficiently complete, and its effects or consequences sufficiently clear. In the interviews, participants described two 'critical incidents' of Web information seeking and use in reply to the following question:

"Please try to recall a recent instance in which you found important information on the Web, information that led to some significant action or decision. Would you please describe that incident for me in enough detail so that I can visualize the situation?"

Where appropriate, participants were prompted with the names of Web sites that were indicated in their WebTracker log files. Besides 'critical incidents,' participants were also invited to comment more broadly on their use of the Web, including their general Web-use strategies and preferences, as well as what they perceived to be positive and negative aspects of Web use.

4.3 Data Analysis

Stage 1: Categorizing Information Seeking Modes

Data analysis proceeded in two stages. In Stage 1, significant episodes of Web-based information seeking were identified from the personal interview transcripts as well as the WebTracker logs. During interviews, participants were asked to recall "critical incidents" or significant episodes of finding and using information on the Web. By reading the transcipts, each episode was analyzed according to its information need, amount of effort, number of web sources consulted, and information use. Based on this analysis, an episode would be categorized as one of the four modes of scanning (undirected/conditioned viewing; informal/formal searching). WebTracker logs were also examined to identify additional significant episodes. Two criteria were used to select episodes: the episode consumed a substantial amount of time and effort; or the episode was a frequently or regularly repeated activity.

Out of a total of 61 episodes identified, 12 were categorized as **undirected viewing**. The most common example of undirected viewing consisted of visits to general news websites such as those of NewsEdge, news.com, and newspapers. In the words of one participant, the goal was to "keep up with what's happening in the world." General news sites acted as gateways to information covering many different subject areas, and provided an efficient way of surveying current developments without a specific information need in mind. Other channels of undirected viewing included portal sites such as CANOE, and large magazine sites such as ZDnet.

Eighteen episodes were categorized as **conditioned viewing**. The most common examples were regular return visits to bookmarked sites, and starting from a particular page that contained links to sites of interest. Thus, a number of participants regularly visited the websites of Microsoft, Novell, and Sun Microsystems in order to monitor new content in selected sections. One participant regularly visited the Novell site for information on upcoming training courses, seminars, and software updates. Another returned to Sun's Java home page periodically to follow developments in the Electronic Commerce Framework and E-commerce tools. A third person habitually scanned the Canada Newswire Site to view press releases from the Federal and Provincial governments. Yet another customized his start-page at MSN with his own topic headings and keywords.

Twenty three episodes were categorized as **informal search**, and these constituted the largest group. The most common examples of informal search were when participants made use of specific query terms such as names of companies, products or technologies to perform simple searches on easily accessible search engines. There were several examples of selecting search engines that were local to a specific site (e.g. a search engine maintained by a company that only indexed its own web pages). Thus, two participants used the local search engine on the website of Forrester Research (a market research firm) to retrieve information about specific companies; another participant used the search engine at the Environmental Protection Agency to retrieve information on ventilation-heating systems for school buildings. Several of the informal searches used well-known search services from Yahoo and AltaVista.

Eight episodes were categorized as **formal search**. Here, participants were intending to use the information formally (e.g. to write policy or planning documents, to provide definitions). Three formal searches utilized several search engines, including meta search services. Two searches attempted to be exhaustively comprehensive: one used four meta search engines to locate a good example of an action plan that could be formally presented to a manager; the other used the DejaNews search engine to

retrieve two author profiles and scan all their postings. Another search was carried out over four days, retrieving high quality resources on Women Advocacy to be included on an institutional site for International Women and Human Rights.

Stage 2: Analyzing Information Seeking Moves
For each of the significant information seeking episodes categorized in Stage 1, the corresponding sections of the WebTracker log were analyzed to determine the browser-based actions that best characterized each episode.

The WebTracker log files were tabulated into large spreadsheets with entries arranged in chronological sequence. Each entry contained a date-time value, followed by a URL or a browser menu action name. Thus it was possible to examine the information seeking **moves** in chronological order in each of the 61 episodes. Data about the sequence of site visits, repetitions of these sequences, movements backwards and forwards between pages, the use of bookmarking, the selection of sites from stored bookmarks, the use of search engines, printing, and other actions and events captured by the WebTracker were examined to trace the selection and development of information seeking moves over the duration of each episode. Using the criteria presented earlier (based on Ellis' model) and summarized in Figure 2, information seeking moves were analyzed to infer whether moves may be categorized as starting, chaining, browsing, differentiating, monitoring, or extracting.

The most common examples of **starting moves** took the form of participants starting their Web sessions from (1) jumpsites that contained links of interest; (2) portal sites; and (3) Intranet entry pages of their organizations. **Chaining moves** occurred when participants followed links from the starting page or some other page. Chaining could be in either direction (backward/forward). **Browsing moves** occurred when participants looked through top-level pages, examined lists of headings, or viewed sitemaps. **Differentiating moves** were when participants bookmarked a page, printed it, or copied its contents. Another indication of differentiating was when a person went direcly to a specific site of known content (e.g. the Microsoft site) by entering its URL. **Monitoring moves** were when participants revisited favorite sites (that have for example been bookmarked or entered into a customized list/page). Although this was uncommon, another indication would be when participants signed up for email or alert services that informed them of new content on the monitored pages. **Extracting moves** were characterized by participants systematically working through a website to extract information of interest. A common method of extracting was to use local search engines that indexed material at their parent sites.

4.4 Results and Discussion

Sixty-one episodes of 'significant' information seeking were identified and categorized according to the framework developed in Section 3. The majority of the episodes were classified as informal search (23) and conditioned viewing modes (18). A smaller number of episodes were undirected viewing (12) and formal search (8). Figure 4 below shows the distribution of the episodes over the four modes of viewing and searching.

The episodes in each mode were examined in terms of their Web moves. In the **undirected viewing** episodes, data collected by the WebTracker application indicated that the most frequently occurring moves were starting and chaining. Thus, participants began at favorite starting pages (news or portal sites) and followed links that they found interesting on those pages. This was usually characterized by a certain amount of movement back and forth using the starting page as anchor.

In the **conditioned viewing** episodes, the most frequently occurring moves were differentiating, browsing, and monitoring. Thus, participants selected a bookmarked page/site, or entered the URL of a site they remembered (differentiating). Another example of differentiating was when participants printed useful pages for their own files or to show to others. These sites/pages were then examined to locate new content of interest (browsing). The most important characteristic of conditioned viewing was that participants regularly or frequently returned to their selected or differentiated sites/pages to check for new information (monitoring).

Information Seeking on the Web

(34 participants, 7 companies, 61 episodes)

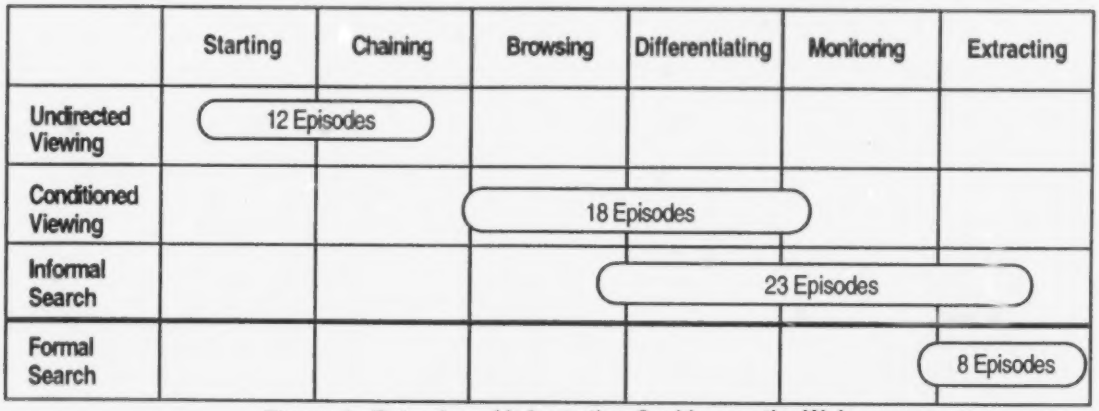

Figure 4. Episodes of Information Seeking on the Web

In the **informal search** episodes, the most frequently occurring moves were differentiating and localized extracting. Thus, participants went directly to selected sites where they expected that the searching they intend to do would likely yield results, e.g. going to a market research firm's site to search for company data, or to a software vendor site to search for software patches (differentiating). Searching at these sites would make use of the local search engines that were dedicated to retrieving information from those sites (localized extracting). Some participants frequently returned to specific sites to perform their informal searches (monitoring).

In the **formal search** episodes, the most frequently occurring move was a relatively intensive and thorough form of extracting, compared with the localized extracting that characterized informal searching. Thus, participants systematically worked through a number of search engines or meta search engines so as to find (all) important information about a topic or item. Formal searches often involved the use of search engines known for their comprehensive coverage, currency, or the inclusion of historical data. The model presented in Section 3.3 and Figure 3 suggested that monitoring would be part of formal searching. *However, for this group of participants, there were no explicit instances of monitoring to support extracting.*

The distribution of information seeking episodes shown in Figure 4 suggests that people who use the Web as part of their work engage in four complementary modes of information seeking as proposed earlier (Figure 1). Each mode is set apart by its information needs, information seeking scope and effort, and the purpose of information use.

Moreover, each mode of information seeking was characterized by information seeking moves that were revealed through recurrent sequences of participants' use of browser functions and features. Undirected

viewing was mainly characterized by starting and chaining; conditioned viewing by differentiating, browsing, and monitoring; informal search by differentiating, and localized extracting; and formal search by systematic, thorough extracting.

The study also introduces an experimental method to operationalize and measure the six patterns of information seeking behaviors identified by Ellis (1989, 1993, 1997) as browser-based actions and events. Recurrent patterns of these actions would indicate that a user is engaging in a particular mode of viewing or searching on the Web. For example, repeated sequences of starting and chaining might suggest undirected viewing (moving back and forth visiting links on a starting page); while sequences of differentiating and extracting might suggest informal search (going to a bookmarked site and doing a local search). Each viewing/searching mode also implies different information needs and information-use goals.

Two other observations can be made. The first concerns "Monitoring," which is keeping up in an area by regularly following particular core or important sources. Two forms of monitoring are possible on the Web: "pull" monitoring is when a user selects a bookmark or enters a URL to revisit a site; "push" monitoring is when a user automatically receives alerts that a monitored site has been updated. Common methods of push monitoring on the Web include subscribing to email newsletters or alerts from the monitored site; setting up a personalized profile or channel; and subscribing to services that track content changes on selected sites. Although most participants in this study would be considered as being Web-savvy, very few of the participants made use of push monitoring techniques: one did use an email alert service; three others tried out a push service, but only for a limited time.

The second observation concerns "Extracting." Extracting on the Web is systematically searching through one or more sites in order to locate information of interest at those sites. In this study, most episodes of extracting employed basic searching strategies. For the most part, search formulations were relatively simple, with advanced features such as Boolean operators, and word truncation or proximity operators rarely utilized. This was the case even when participants appeared to be working in the formal search mode. There were no instances of participants accessing search-engine help instruction pages to improve their searches.

5 SUMMARY

The research presented here suggests that people who use the Web as an information resource to support their daily work activities engage in a range of complementary modes of information seeking, varying from undirected viewing that does not pursue a specific information need, to formal searching that retrieves focused information for action or decision making. Each mode of information seeking on the Web is distinguished by the nature of information needs, information seeking tactics, and the purpose of information use. The information seeking tactics characterizing each mode were revealed by recurrent sequences of browser actions initiated by the information seeker. Thus, undirected viewing is characterized by starting and chaining actions; conditioned viewing is characterized by differentiating, browsing, and monitoring actions; informal search is characterized by differentiating and localized extracting; and formal search consisted of systematic, thorough extracting.

Overall, the study suggests that a behavioral framework that relates motivations (the strategies and reasons for viewing and searching) and moves (the tactics used to find and use information) may be helpful in analysing Web-based information seeking. The study also suggests that multiple, complementary methods of collecting qualitative and quantitative data may be integrated within a single study to compose a more nuanced portrayal of how individuals seek and use Web-based information in their natural work settings.

(This research is supported by a grant from the Social Sciences and Humanities Research Council of Canada. The WebTracker application was developed by Ross Barclay, a master's student at the Faculty of Information Studies, University of Toronto. More information about the project is at http://choo.fis.utoronto.ca/esproject/)

References

Aguilar, Francis J. 1967. *Scanning the Business Environment.* New York, NY: Macmillan Co.

Aguilar, Francis J. 1988. *General Managers in Action.* New York, NY: Oxford University Press.

Catledge, L. D. and J. E. Pitkow. 1995. Characterizing Browsing Strategies in the World Wide Web. *http://www.igd.fhg.de/www/www95/papers/80/userpatterns/UserPatterns.Paper4.formatted.html*

Chang, Shan-Ju and Ronald E. Rice. 1993. Browsing: A Multidimensional Framework. In *Annual Review of Information Science and Technology* , ed. Martha E. Williams. Medford, NJ: Learned Information.

Choo, Chun Wei. 1998. *Information Management for the Intelligent Organization: The Art of Scanning the Environment.* Second ed. Medford, NJ: Information Today, Inc.

Choo, Chun Wei, Brian Detlor, and Don Turnbull. 1998. A Behavioral Model of Information Seeking on the Web: Preliminary Results of a Study of How Managers and IT Specialists Use the Web. In Proceedings of *61st ASIS Annual Meeting* held in Pittsburgh, PA, edited by Cecilia M. Preston, Vol. 35: 290-302. Information Today Inc.

Daft, Richard L. and Karl E. Weick. 1984. Toward a Model of Organizations as Interpretation Systems. *Academy of Management Review* 9, no. 2: 284-295.

Ellis, David and Merete Haugan. 1997. Modelling the Information Seeking Patterns of Engineers and Research Scientists in an Industrial Environment. *Journal of Documentation* 53, no. 4: 384-403.

Ellis, David, D. Cox, and K. Hall. 1993. A Comparison of the Information Seeking Patterns of Researchers in the Physical and Social Sciences. *Journal of Documentation* 49, no. 4: 356-369.

Ellis, David. 1989. A Behavioural Model for Information Retrieval System Design. *Journal of Information Science* 15, no. 4/5: 237-247.

Flanagan, John C. 1954. The Critical Incident Technique. *Psychological Bulletin* 51, no. 4: 327-358.

Huberman, Bernardo A., Peter L. Pirolli, James E. Pitkow, and Rajan M. Lukose. 1998. Strong Regularities in World Wide Web Surfing. *Science* 280, no. 5360: 94-97.

Marchionini, Gary M. 1995. *Information Seeking in Electronic Environments.* Cambridge, UK: Cambridge University Press.

Tauscher, Linda and Saul Greenberg. 1997. How People Revisit Web Pages: Empirical Findings and Implications for the Design of History Systems. *International Journal of Human-Computer Studies* 47: 97-137.

Tauscher, Linda and Saul Greenberg. 1997. Revisitation Patterns in World Wide Web Navigation. In Proceedings of *CHI 97 Human Factors in Computing Systems* held in Atlanta, Georgia, USA, edited by Steven Pemberton, Vol. 399-406. CHI 97.

Weick, Karl E. and Richard L. Daft. 1983. The Effectiveness of Interpretation Systems. In *Organizational Effectiveness: A Comparison of Multiple Models*, ed. Kim S. Cameron and David A. Whetten, 71-93. New York, NY: Academic Press.

Wilson, T. D. 1997. Information Behaviour: An Interdisciplinary Perspective. *Information Processing & Management* 33, no. 4: 551-572.